MAS223 Statistical Modelling and Inference Chapter 4: Case Studies

Dr Nic Freeman

School of Mathematics and Statistics, University of Sheffield

Autumn Semester, 2015

Animal study

A vet proposes a new treatment protocol for a certain animal disease. With current methods about 40% of animals with this condition survive six months after diagnosis.

After one year of using the new protocol, 15 animals have been treated and followed to six months after diagnosis, of whom 6 survived.

After two years a further 55 animals have been followed to the six months mark, of whom 28 survived. So in total we have 34 animals surviving out of 70.

Model

We model the data as independent Bernoulli trials, giving the binomial observation $X =$ number of animals surviving, and $X \sim Bi(70, \theta)$, where θ is the probability of survival with the new treatment.

We have actually observed $x = 34$.

Interest in this case study clearly rests directly on whether the new protocol increases the six-month survival rate, i.e. whether $\theta > 0.4$.

Calculations

Example 39: Likelihood calculations for animal study

Cost effectiveness

An area of growing importance is that of judging the cost-effectiveness of medical treatments. Very expensive treatments cannot be justified even when they are very effective, since there is only a finite total resource to devote to health care.

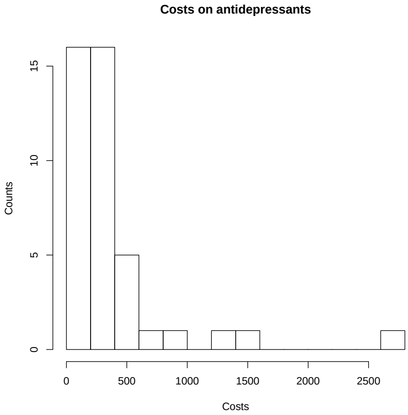
Data

These data concern the comparison of two treatments for depression (42 on Antidepressants and 39 on Counselling) for efficacy and for cost.

There are data on 81 patients, 42 on Antidepressants and 39 on Counselling; outcome ‘good’ or ‘not good’ (after a year); costs (of treatment) over the year are in £.

The interest will be in comparing mean costs: hence the question is how to estimate this well. For simplicity, we will focus on the costs for those on antidepressants only.

Histogram of Antidepressant costs

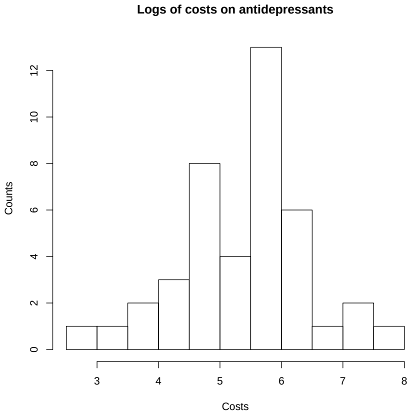


A model for those on antidepressants

Based on the histograms, it would not be reasonable to suppose that the costs followed a normal distribution.

For skewed strictly positive data, such as costs, working with the log of the original values is often a good idea. In this case (see next slide) their histogram looks roughly bell-shaped.

Histogram of Antidepressant log-costs



A model for those on antidepressants

So, to model this situation we will let $x_i = \log c_i$, ($i = 1, 2, \ldots, 42$), and suppose that the x_is are independent observations from a normal distribution.

We therefore assume $X_i \sim N(\mu, \sigma^2)$ – so there are two parameters, μ and σ^2. [We say that $C_i = \exp X_i$ has a **lognormal distribution**, $C_i \sim lN(\mu, \sigma^2)$.]

Calculations

Example 40: Likelihood calculations for antidepressant costs

Ecotoxicology

Our next dataset comes from a real problem concerning standards for toxic chemicals in rivers. The purpose of setting a standard is to control the concentrations of pollutants at levels that protect aquatic animals.

However, there are a very large number of species that there is a need to protect, not just fish but also water snails, insects, leeches, etc. Data on toxicity of any given chemical is available for only a small number of the species that are of interest.

Data

In this analysis, all the available toxicity data for the insecticide chlorpyrifos was identified.

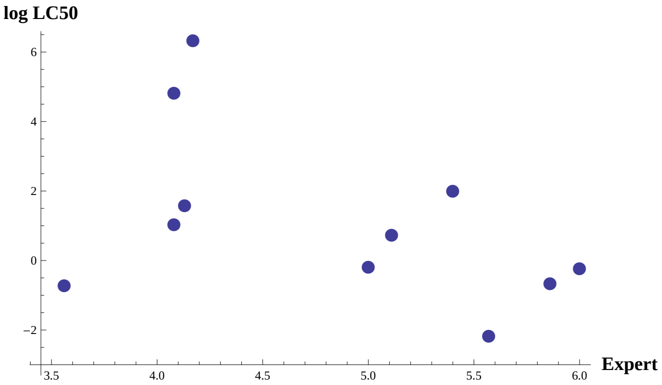
The data are estimates of the LC50 (the concentration that will kill 50% of the individuals) for 96-hour exposure to chlorpyrifos. The 11 items of 96-hour LC50 data found are tabulated. There is a wish to extrapolate from the published data to say something about the wider collection of species that will be present in a river.

Expert assessments

Some freshwater biologists scored the sensitivity to chlorpyrifos of taxonomic groups on a scale from 1 to 8. They were also asked to score their own knowledge of each taxonomic group from 0 to 5. The sensitivity scores for each taxon were then weighted according to expertise.

The idea of getting the expert assessments was to try to link these to the toxicity data, in order to predict toxicity in other species. It is not practicable to ask the experts about all the possible species that might be present in a British river, so they were asked about groups that contain related species (which hopefully will have similar sensitivity to chlorpyrifos).

Toxicity data



Toxicity data. The log of LC50 plotted against expert sensitivity scores appears to support a linear relation between them.

A model

The following statistical model will be assumed for these data.

For $i = 1, 2, \ldots, 11$, let $y_i = \log z_i$, where z_i is the ith toxicity measurement, and let x_i be the corresponding average score of the experts.

Suppose that a linear regression relationship applies between these variables, so that

$$y_i = \alpha + \beta x_i + \epsilon_i \ ,$$

where the ϵ_is are independent $N(0, \sigma^2)$ errors, and $\boldsymbol{\theta} = (\alpha, \beta, \sigma^2)$ are the unknown parameters.

Expectations and calculations

We should expect that β will be negative, because increasing sensitivity to chlorpyrifos should be associated with a lower LC50. The plot of the data provides support to this expectation.

Example 41: Setting up likelihood for linear model